

# Level-1C Product from AIRS: Principal Component Filtering

Evan M. Manning<sup>\*a</sup>, Yibo Jiang<sup>a</sup>, Hartmut H. Aumann<sup>a</sup>, Denis A. Elliott<sup>a</sup>, Scott Hannon<sup>b</sup>

<sup>a</sup>Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA, USA 91109

<sup>b</sup>University of Maryland Baltimore County, Baltimore, MD

## ABSTRACT

The Atmospheric Infrared Sounder (AIRS), launched on the EOS Aqua spacecraft on May 4, 2002, is a grating spectrometer with 2378 channels in the range 3.7 to 15.4 microns. In a grating spectrometer each individual radiance measurement is largely independent of all others. Most measurements are extremely accurate and have very low noise levels. However, some channels exhibit high noise levels or other anomalous behavior, complicating applications needing radiances throughout a band, such as cross-calibration with other instruments and regression retrieval algorithms. The AIRS Level-1C product is similar to Level-1B but with instrument artifacts removed. This paper focuses on the “cleaning” portion of Level-1C, which identifies bad radiance values within spectra and produces substitute radiances using redundant information from other channels. The substitution is done in two passes, first with a simple combination of values from neighboring channels, then with principal components. After results of the substitution are shown, differences between principal component reconstructed values and observed radiances are used to investigate detailed noise characteristics and spatial misalignment in other channels.

**Keywords:** AIRS, Level-1C, Principal Components, NEdT, infrared sounder, grating array, cross-calibration

## 1. INTRODUCTION AND REVIEW

### 1.1 AIRS Instrumentation

The Atmospheric Infrared Sounder (AIRS), launched aboard NASA’s EOS Aqua spacecraft on May 4, 2002, is a grating array spectrometer having 2378 channels sensitive in the range 3.7 to 15.4 microns. The spectral resolution ( $\lambda/\Delta\lambda$ ) is about 1200. A combination of a design philosophy having radiometric accuracy as a foremost goal, cooled and temperature-controlled spectrometer hardware (including most of the optics), and thorough pre-flight calibration have made AIRS a superb instrument that produces very high quality radiance data<sup>1</sup>. AIRS will have completed ten years in routine operations at the end of August 2012. The instrument remains healthy and it is hoped that it will continue to operate and produce high quality data for several more years. Such a long data record makes AIRS an excellent candidate for producing useful records for climate trend analyses. To maximize its utility as a source of climate data, any instrument must be capable of being accurately cross-calibrated with other instruments so that lengthy multi-instrument records may be generated.

Because AIRS is a grating array spectrometer, not a Fourier transform spectrometer, each of the 2378 channels is independent of the others, is separately calibrated, and has the potential for noise behavior that is different from the other channels—even neighboring ones. In fact, the noise behavior of a channel can change (or a channel could even stop functioning) suddenly, perhaps due to a radiation hit, without any other channel being affected. Channel variability in noise characteristics, if unaccounted for, complicates error estimates, noise estimates, and cross-instrument calibration.

This document describes the methodology and procedure that produces cleaned-up AIRS spectra which can safely be used for instrument cross-calibration and studies of climate trends. The final output will be a prescription for producing Level 1C (L1C) radiances that will become publicly available in a future version of the AIRS science product generation software (currently in Version 5) operating at the Goddard Earth Sciences Data and Information Services Center (GES DISC). Level 1A (L1A) products contain raw detector counts. Level 1B (L1B) consists of radiometrically calibrated radiances<sup>2</sup>. Both L1A and L1B have been produced since the start of routine instrument operations in September 2002.

---

\* [evan.m.manning@jpl.nasa.gov](mailto:evan.m.manning@jpl.nasa.gov); phone 1 818 354-1172; fax 1 818 393-4918; [jpl.nasa.gov](http://jpl.nasa.gov)

L1C will be cleaned up spectra, in which dead or unusually noisy channels have been marked and their radiances corrected or replaced using information from well-behaved correlated AIRS channels. Other artifacts will also be removed, but are not detailed here. An earlier state of this algorithm has previously been presented<sup>3</sup>.

## 1.2 Level-1C Motivation

An ideal spectrum should have the following properties:

- 1) All channels have only Gaussian noise, so that far outliers do not interfere with integration over the pass band or interpolation during resampling.
- 2) All frequencies are included (at the design resolution) with no gaps and no overlapped spectral regions.
- 3) Adjacent channels are truly independent measurements, with adjacent-channel correlations minimized.
- 4) The channel frequencies are fixed in time.

The AIRS instrument performance deviates from the ideal described above in small ways. Actually, its in-flight performance has far exceeded its specifications, which were developed with temperature and humidity profile measurement in mind. The originally envisioned use of AIRS was for non-climate-related studies of atmospheric phenomena and for improved weather prediction. Because of its exceptional performance it is now being used for climate studies. AIRS climate data records can be improved if the deviations from ideal performance are accounted for.

AIRS, a grating array instrument, has seventeen detector modules spread across the focal plane. The hardware design was simplified by permitting small gaps in the frequency coverage between some modules, and small overlaps in coverage between others. These gaps and overlaps complicate the task of integrating the spectrum over the pass band to enable cross-instrument comparisons and calibrations.

Because of the large frequency range of the entire instrument, each detector module is made from different material having different properties, and unique read-out integrated circuits (ROIC). The shortwave modules include circuits to remove spikes resulting from radiation hits, while the spikes resulting from radiation hits are very obvious in the longwave modules. Some channels exhibit non-Gaussian noise including “pops” (temporary changes in output level) and cold scene noise (scene-dependent noise larger at lower signal levels)<sup>4</sup>. The AIRS channels and their ROIC’s have different susceptibilities to radiation hits and to the slow build up of total radiation dosage throughout the mission. Thus, a channel that has exhibited very low noise for years can suddenly undergo a noise increase. It is also possible for a channel that has been affected by a radiation hit to recover after a period of hours or days. So noise can vary independently over time from channel to channel.

Each channel occupies its own physical space on the focal plane. That implies that slight changes in channel frequency can occur, due primarily to changes in temperature gradients within the spectrometer optical train. The AIRS spectrometer’s temperature is tightly controlled at one location by a servo-controlled heater. But small changes in internal gradients can occur and result in very small channel frequency shifts with time (~0.5% of SRF width per year).

Grating array spectrometers, by their very nature, perform extremely well as far as point #3 above is concerned. However, the independence of the channels means that the other goals are not met automatically, and so must be handled explicitly. In this document, we describe the first step towards a final L1C product that will fulfill the ideal. In this paper, we address point #1 above (variable and/or non-Gaussian noise). Note that, because of the fact that the AIRS spectrometer is temperature controlled, point #4 (channel frequencies vs. time) has only a minor effect on AIRS data quality. The observed instrument frequency stability far exceeds the specification. But when the more stringent requirements for climate records are considered, it becomes clear that frequency shifts do need to be handled eventually.

Once “ideal” spectra have been produced, the use of AIRS data for many studies will be simplified and errors more easily estimated. Most importantly, it will be possible to cross-calibrate AIRS with other climate instruments more easily and reliably than that can be done today.

Figure 1 shows a typical L1B spectrum, where noisy or dead channels are indicated by the red dots at the bottom of the plot. The 17 modules and their location in the spectrum are also labeled in this plot. Of the 2378 AIRS spectral channels, about 100 (5%) have no response (dead) or have more than 2 K noise (compared to the nominal noise of ~0.2K). The presence of these problem channels is overemphasized in Figure 1.

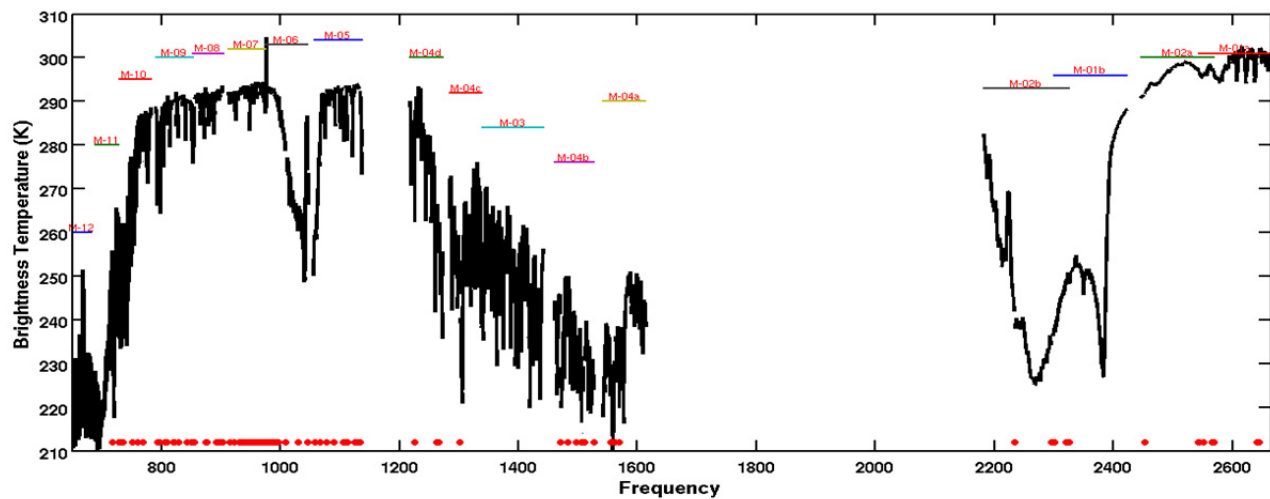


Figure 1. AIRS nighttime spectrum from January 1, 2006. The red dots at the bottom of the plot indicate dead or bad channels.

## 2. SPECTRUM CLEANING

This section describes the spectrum cleaning algorithm and its implementation. Test results are shown and then reconstruction differences are used to explore noise and spatial characteristics of the AIRS Level-1B product. The major steps in the spectrum cleaning process are:

- Identification of obviously bad channels.
- First-order replacement of these bad channels using the buddy system. This step is required or else the PCA diverges.
- PCA reconstruction of the spectrum and replacement of the bad channels.
- Scan the spectrum for outliers based on PCA reconstruction and identify noise spikes, which can also be removed.

### 2.1 Dead and noisy channel detection

For each observation, typically about 110 of 2378 channels are marked for replacement and others are excluded from use in replacing other channels. This process generally relies on noise levels, error flags, and radiances from the Level-1B product, so in theory different channels could be selected for replacement in each granule, but in operations we see mostly the same channels, including ~50 permanently dead or noisy from launch through the first 10 years of operations. An additional seven channels are marked permanently bad because they have been unusable since the launch of the AIRS instrument but are not always flagged as bad in the Level-1B product. These include 5 channels that are cross-wired, so they actually have low noise, but they do not observe the spectral region they were intended to.

Channel replacement is required whenever:

- 1) Noise Equivalent delta Temperature (NEdT) at 250 K scene temperature exceeds a threshold (currently 2.0 K) or NEdT is negative (indicating noise could not be characterized).
- 2) L1B provides no radiance value.
- 3) L1B CalFlag indicates a problem with gain or offset calculations, bad telemetry, or a “pop” event. (Only pop events are seen when the instrument is in operational mode.)
- 4) Channels are on a list of permanently bad detectors.

Channels are not replaced but are excluded from use in replacing other channels whenever:

- 1) NEdT at 250 K exceeds 1.0 K.
- 2) Observed radiance is negative.

- 3) AB\_State from the current channel properties file marks the channel lower quality with state  $> 2$ .

The median NEdT at the reference temperature of 250 K is 0.2 K. The further detection of "bad" or "noisy" channels with the help of principal component analysis will be discussed later.

## 2.2 The Night-Time Clear-Sky Training Set

While the AIRS spectrum consists of 2378 independent spectral channels, the information content of the spectrum as a whole is much less. For example, about 700 of the 2378 channels have weighting functions that peak at the surface. For all AIRS channels, there exist other AIRS channels that are radiometrically similar. We can therefore replace the radiance of known bad or noisy channels with ones computed using these approximately equivalent channels or "buddies".

We prefer to select buddies for each channel by finding channels with similar statistics over a representative set of observed spectra. However, for our 7 permanently bad channels and about 100 others that are dynamically bad on the day used for this set, we can't do this. So these channels are first replaced using synthetic (simulated) spectra, then we can produce a training set with all channels good for training the final buddy replacement list and the principal component step.

The first training set is derived from AIRS night-time clear-sky spectra model simulation by UMBC radiation transfer model with 49 climatology spectra at satellite zenith angles 0, 10, 20, 30, 40, 50 degrees. Some of the spectra are shown in Figure 2.

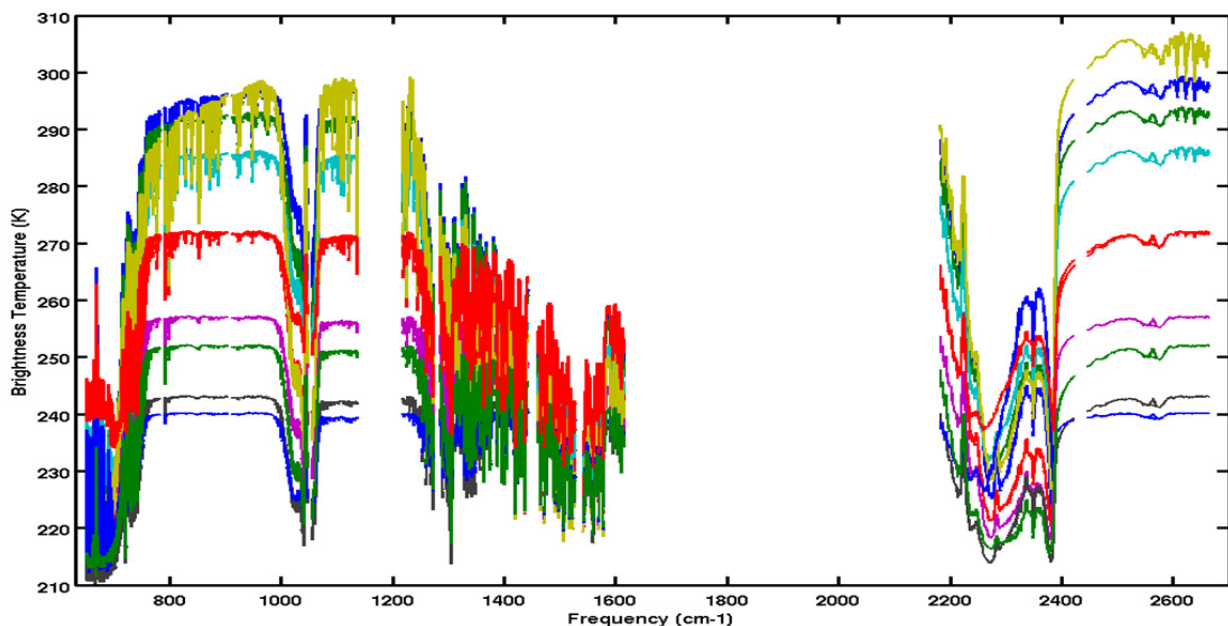


Figure 2. Examples of night-time clear-sky training sets.

## 2.3 The global training set

The second training set (Figure 3) uses real AIRS spectra observed on January 1, 2008. There are 21,502 profiles in this training set, and they include both day-time and night-time spectra, and the spectra naturally include variability in surface types, cloud types, and dust. This was the largest number of samples in the training set that the PCA training algorithm could accommodate. The spectra were selected such as to represent all the factors that may affect the measurement. Each of the 21,502 spectra was cleaned and filled by the first set "buddy" replacement

process. Some of the spectra are shown in Figure 3. This training set was used to find the new set of “buddies” which will be used in LIC process, and create the eigenvectors used for the detection and replacement of bad-channels by the Principal Component Analysis method<sup>5</sup>. It was originally thought that any day would be equally good, because winter, summer, etc. conditions all will be encountered somewhere on the globe. But January 1<sup>st</sup> is near the start of northern winter, so this training set lacks the coldest cases, such as those seen in Antarctica in the coldest parts of southern winter. Evidence of this is discussed in section 2.6 and shown in Figures 8-10.

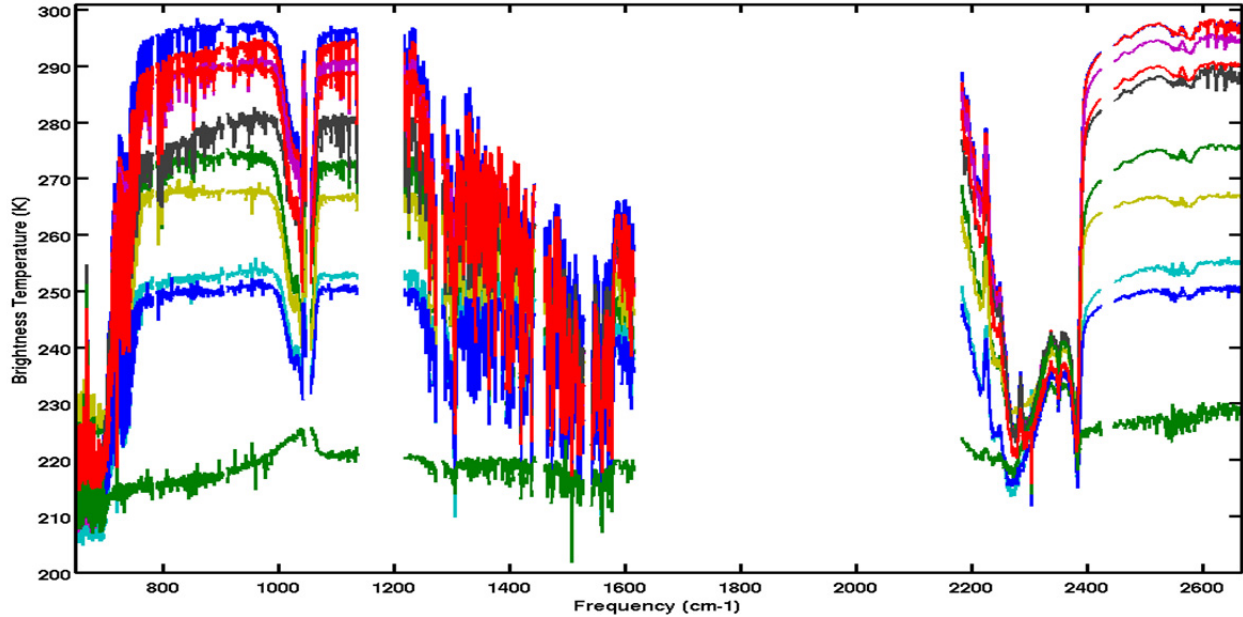


Figure 3. Example spectra from global training set.

The first step in selection of profiles for inclusion in the training set was to evaluate the median brightness temperature (BT) for each of the 17 detector modules for each spectrum. Then the median of those medians was calculated as the overall median BT for the spectrum. The key parameters for distinguishing profiles were:

- 1) Overall median BT
- 2) Delta BT between module M-12 (649-682 cm<sup>-1</sup>) BT and overall median BT
- 3) Delta BT between module M-01b (2300-2422 cm<sup>-1</sup>) BT and overall median BT
- 4) Delta BT between module M-01a (2546-2665 cm<sup>-1</sup>) BT and overall median BT
- 5) Delta BT between module M-02a (2446-2564 cm<sup>-1</sup>) BT and overall median BT
- 6) Delta BT between module M-04d (1217-1272 cm<sup>-1</sup>) BT and overall median BT

6-dimensional bins were created by dividing each parameter into intervals:

- 1) overall median BT: 2K bins
- 2-5) delta M-12 through delta M-02a: range of values encountered is divided into 12 equal bins
- 7) delta M-04d: range of values encountered is divided into 8 equal bins

Then from each non-empty bin, one spectrum was selected at random. Assuming the overall median BT covers a span of about 100 K, this gives  $50 \times 12 \times 12 \times 12 \times 12 \times 8 \approx 8$  million bins, but most combinations are empty.

## 2.4 Buddy system algorithm

In order to fill in the bad channels with reasonable values (“buddies”) from the same spectrum, the most correlated channels (minimum standard deviation) are found based on the training set. The BT of the bad channel is then replaced by the BT from the most correlated channels. The correlated channel replacement list is calculated based on the following formula which calculates the averaged deviation from the channel to be filled. Note that the “buddy” channels always come from the same detector module (Figure 1) as the channel that is being replaced.

$$\delta T(k, j) = \sqrt{\frac{1}{n} \sum_i^n (T_{ij} - T_{ik})^2}$$

$n$ : number of spectra in the training set  
 $i$ : spectrum index, range from 1 to  $n$   
 $j$ : channel number range from 1 to 2378  
 $k$ : channel to be filled range from 1 to 2378

$T$  represents the brightness temperature in the above formula,  $i$  is the spectrum index in the training set,  $j$  is the replacement channel number and  $k$  is the channel number to be filled.  $\delta T(k, j)$  represents the averaged deviation of each individual channel  $j$  to the filled channel  $k$ . For each channel  $k$ ,  $\delta T(k, j)$  values are sorted in the ascending order and the first 100  $j$ s with the least deviations from channel  $k$  are selected and will be used to replace or fill the bad channels. This process is repeated for 10 15-K scene brightness temperature ranges with limits of {220, 235, 250, 265, 280, 295, 310, 325, 340, 355, 370} K. The resulting 100  $j$ s (integer array size 2378x100) and the associated deviations (double array size 2378x100) and biases are stored in a lookup table which will be used in the later process.

The brightness temperature of the channel to be filled or replaced then is the average of the four best correlated of the useable channels weighted by the deviations from the filled or replaced channel. In L1C implementation, the four most correlated channels are used in the final calculation of the brightness temperature  $T_k$  of the replaced channels.

$$T_k = \frac{\sum_j (T_j + fB_r(k, j)) \frac{1}{\delta T_r(k, j)}}{\sum_j \frac{1}{\delta T_r(k, j)}}$$

$j$ : channel number range from 1 to 2378  
 $k$ : channel to be filled range from 1 to 2378  
 $r$ : brightness temperature range from 1 to 10  
 $B$ : brightness temperature bias  
 $f$ : bias scale factor

The use of a bias allows us to find much better matches than otherwise, but there is a catch. The degree of bias between neighboring channels is highly scene dependent: the more clouds, the less spectral contrast. Therefore, for each channel to be replaced and in each spectrum, we first determine a bias scale factor  $f$ .  $f$  is selected from among 9 possible values {0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00} by finding the value that minimizes the penalized standard deviation of the candidate fill values  $T_j + fB_r(k, j)$ . The penalty function requires that the standard deviation be 4x smaller for the extreme bias scale factors of 0.00 and 2.00 than for the nominal value of 1.00, which makes the process favor using the nominal value.

Figure 4 shows a typical AIRS spectrum after the “buddy-system” replacement of dead or noisy channels. All visible outliers are suppressed and reasonably good values are provided for missing channels. Errors of 1-5 K are still important but not obvious on this scale, so we demonstrate the further improvements made by PCA in Figure 6 below and then characterize the final cleaned product in section 2.6.

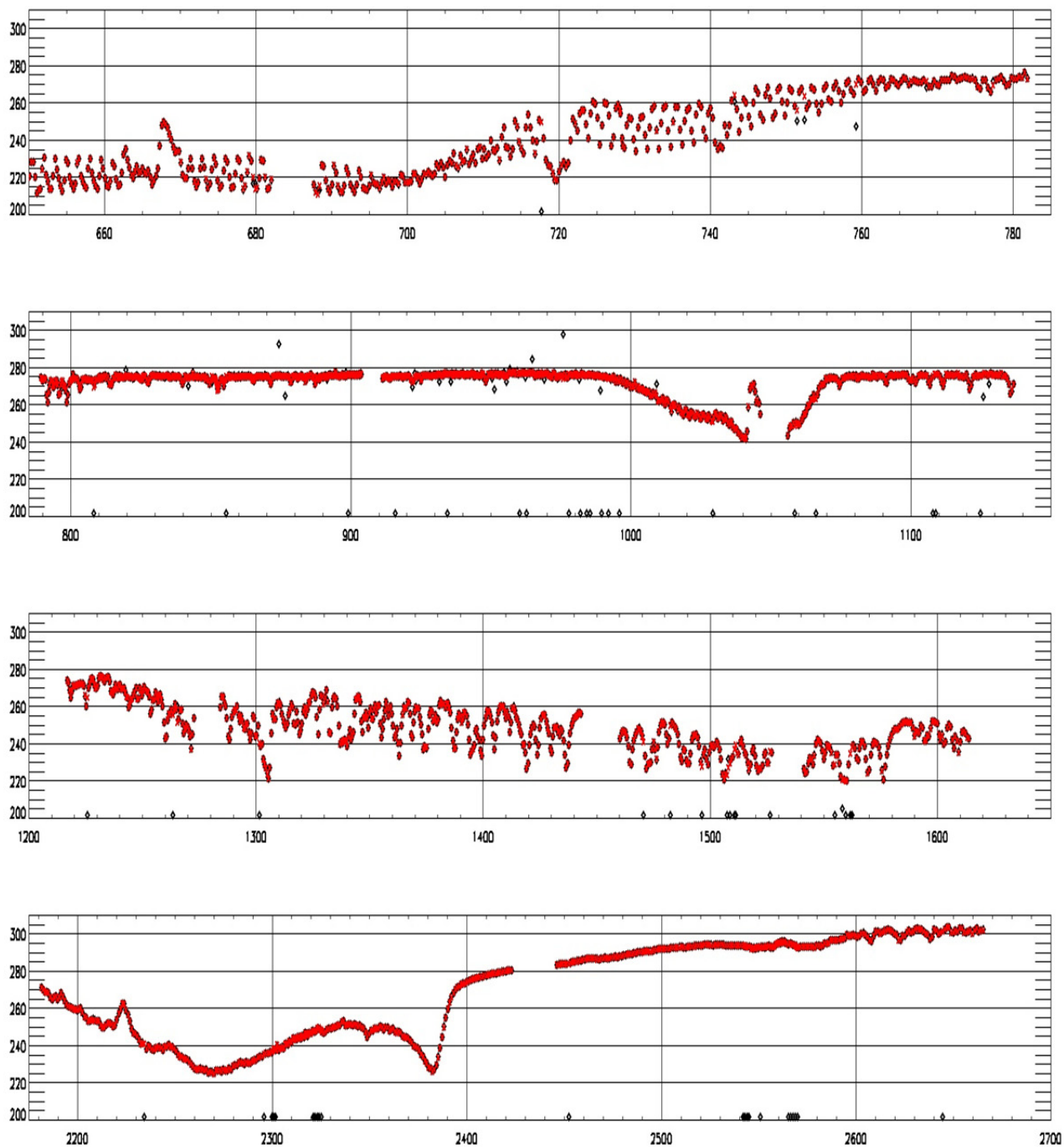


Figure 4. A typical AIRS spectrum after the first pass “buddy-system” replacement of dead or noisy channels. The black diamond symbol represents the original L1B spectrum and the red circle is the cleaned spectrum. Diamonds along the bottom of the spectrum are channels for which the L1B value was out of range or L1B did not provide a value.



## 2.5 Principal component analysis and AIRS spectrum reconstruction

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. PCA is a simple, non-parametric method of extracting relevant information from complicated data sets. Any data can be expressed as linear combinations of eigenvectors, where the first principal component (associated with the first eigenvector) accounts for most of the variability in the data, and each succeeding component accounts for remaining variability. The real spectrum training set is used to create the fixed set of eigenvectors. The first four major principal components and their eigenvalues of the training set are shown in Figure 5. The eigenvalues fall sharply, and the first twenty components account for over 99% of the signal. The number of components used to reconstruct the spectra increases, the reconstructed spectra will increasingly match the observed spectra, eventually including noise artifacts. Since we use the PCA reconstructed spectrum for noise detection, we choose 400 components in our PCA reconstruction scheme.

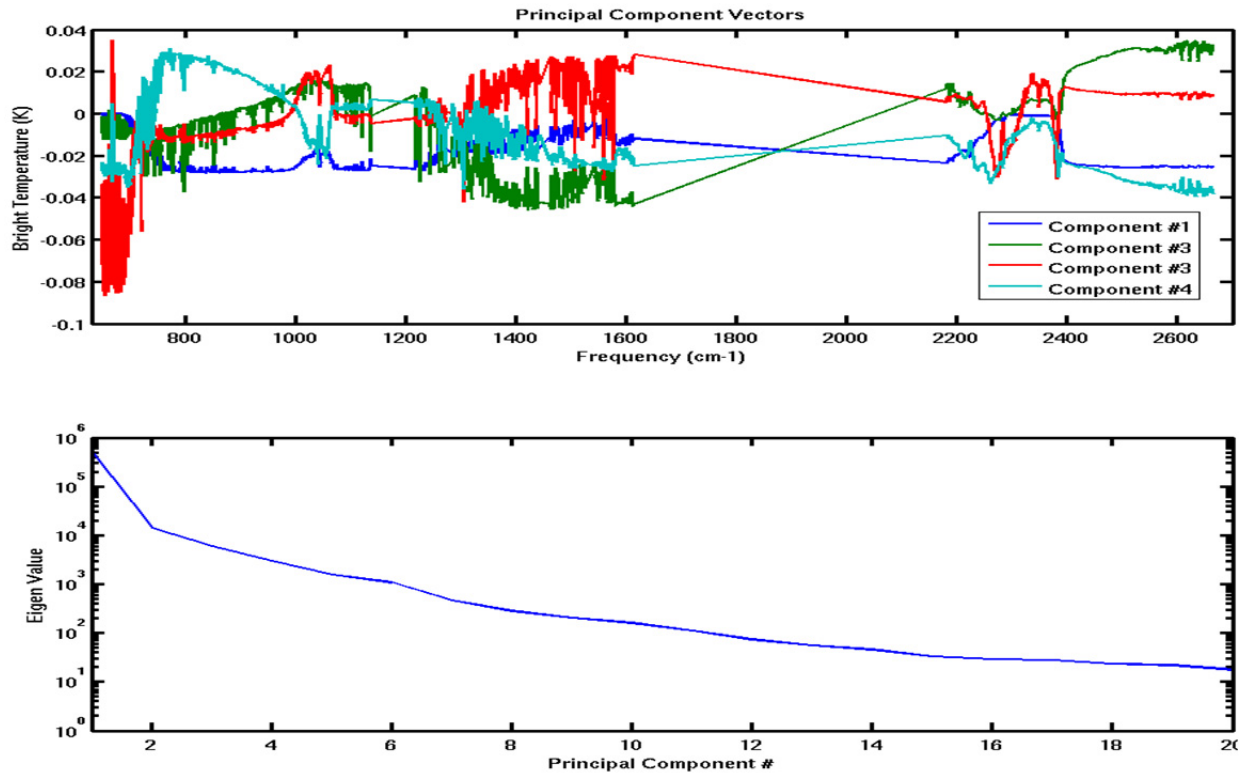


Figure 5. Principal component analysis of the training set for the first four components (top panel) and the eigenvalues for the first twenty components (bottom panel).

The observed spectrum  $T$  can then be reconstructed from the PCA using  $T = E^T T_0 + \bar{T}_T$  where  $T$  represents the reconstructed brightness temperature,  $E^T$  is the transposed eigenvector,  $T_0$  is the original spectrum, and  $\bar{T}_T$  represents the average of the training set.

Once a spectrum has been cleaned by the “buddy-system” replacement, the principal component analysis is used in order to further detect and correct the cleaned spectrum.

At the 100+ K scale of Figure 4, even the buddy system replacement looks perfect, but by zooming in we can observe the additional accuracy provided by the PCA step. Figure 6 shows a close-up of the 2290-2330  $\text{cm}^{-1}$  spectral region of the same spectrum shown in Figure 4. Now in addition to black diamonds for the raw L1B spectrum and red Xs for the buddy replacements we have green +s for the PCA replacement. All good channels in L1B are not replaced and so have the same values for all 3 sets. 2300-2326  $\text{cm}^{-1}$  is an “overlap” region where channels from two detector modules



observe almost the same frequencies and also where there are several bad channels. The long red and green line segments connect the ends of the modules and make it easy to see the extent of the overlap region. Several channels at the ends of the modules ( $2300\text{--}2302\text{ cm}^{-1}$  and  $2321\text{--}2325\text{ cm}^{-1}$ ) are marked bad and replaced. For all of these channels the green PCA replacement values are more similar to values from good channels in the other band than are the red buddy-system values. There is also one odd point at  $2302.5\text{ cm}^{-1}$  where the two replacement values agree quite closely but disagree with the observation from the other band by over 1 K. The specific cause of this is not known, but this magnitude of error is within expectations.

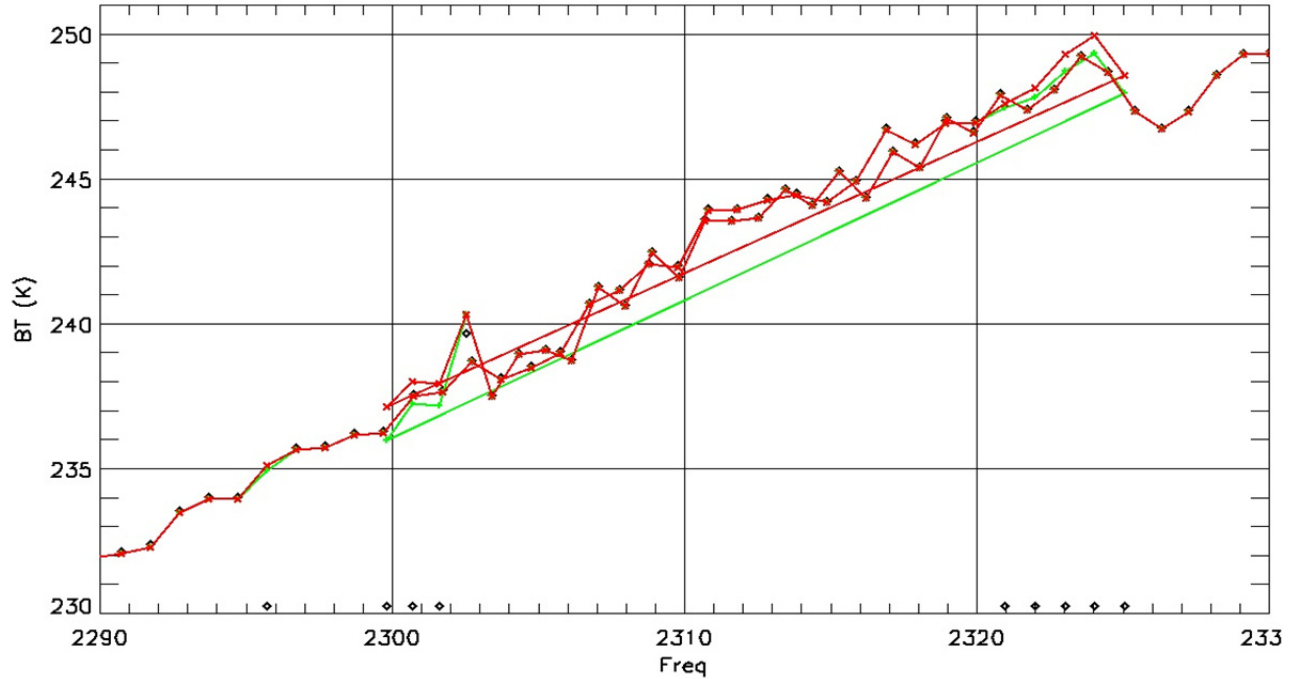


Figure 6. Close-up of buddy system and PCA replacement values in an overlap region. Black diamonds: L1B. Red Xs: Buddy system. Green +s: PCA.

## 2.6 Algorithm Validation

AIRS spectra themselves are the great source for the validation of the spectrum cleaning algorithm. For this test we ran all 240 granules from one day, 2007-09-06 ten times, each time knocking out every 10<sup>th</sup> channel from a different starting point. Then all of the replacement values were combined into a single spectrum for each observation. This spectrum is then compared to the actual observed Level-1B spectrum in order to evaluate the filling algorithm. This is a real stress test, because where approximately 5% of channels are normally filled because of static or dynamic problems, now 15% of channels are replaced at once. Results shown include only channels with reported noise level  $\leq 0.6\text{ K}$  and only scenes where the L1B brightness temperature is at least 220 K.

Figure 7 shows the bias (black) and standard deviation (blue) by channel over the full day of data. The results are excellent, with bias almost always within  $\pm 0.1\text{ K}$  and std dev rarely much larger than the noise level. Even the outliers have bias under 1 K and std dev under 1.5 K, very good for the purposes of L1C.

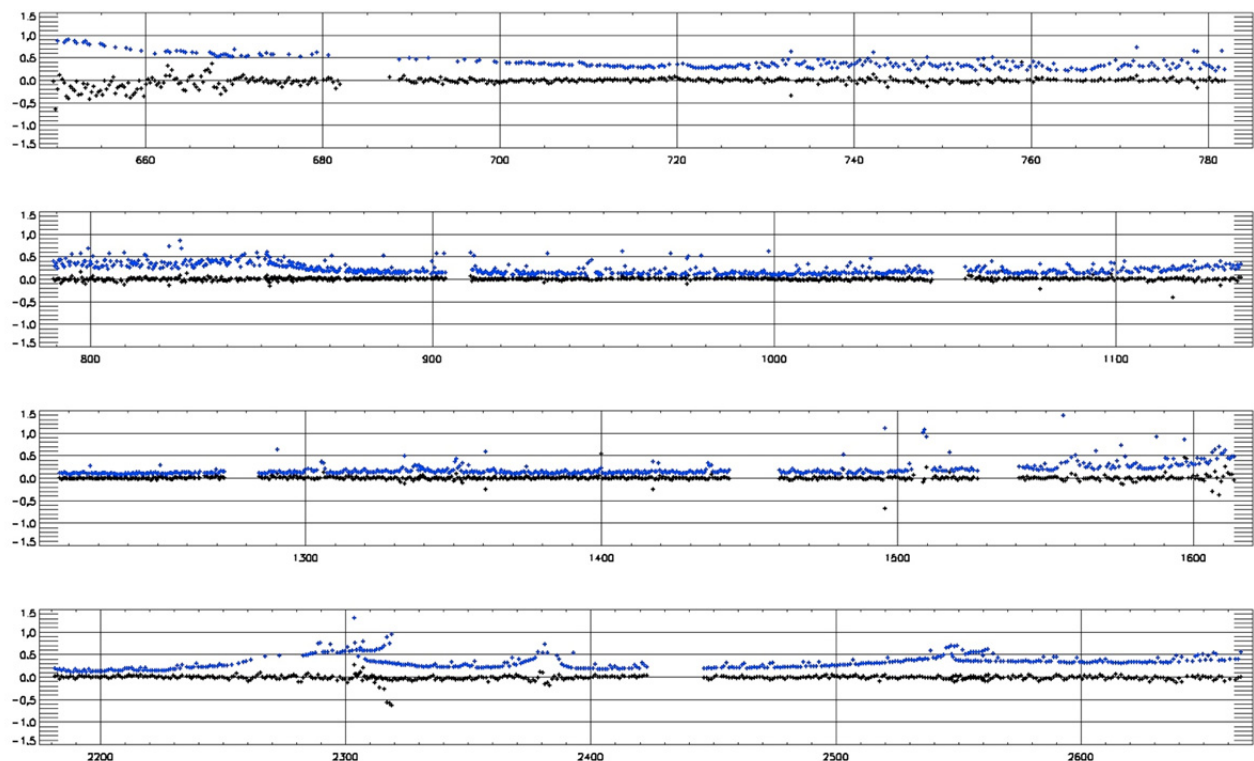


Figure 7. Bias (black) and standard deviation of reconstruction error (K) as a function of frequency ( $\text{cm}^{-1}$ )

We can investigate this further by dividing the globe into 6 latitude zones. Figures 8 (bias) and 9 (standard deviation) give statistics by zone. From about  $649\text{--}700\text{ cm}^{-1}$  we see a cold bias in the tropics, and throughout the spectrum but especially in the shortwave region from  $2300\text{--}2660\text{ cm}^{-1}$  we see enhanced noise in the Antarctic. The enhanced noise may be caused by a lack of austral winter cases in the training set or might just be an artifact that noise levels in brightness temperature units get larger as scenes get colder.

In any case, the results are still excellent with all biases within  $\pm 1.5\text{ K}$  and standard deviation almost always in this range.

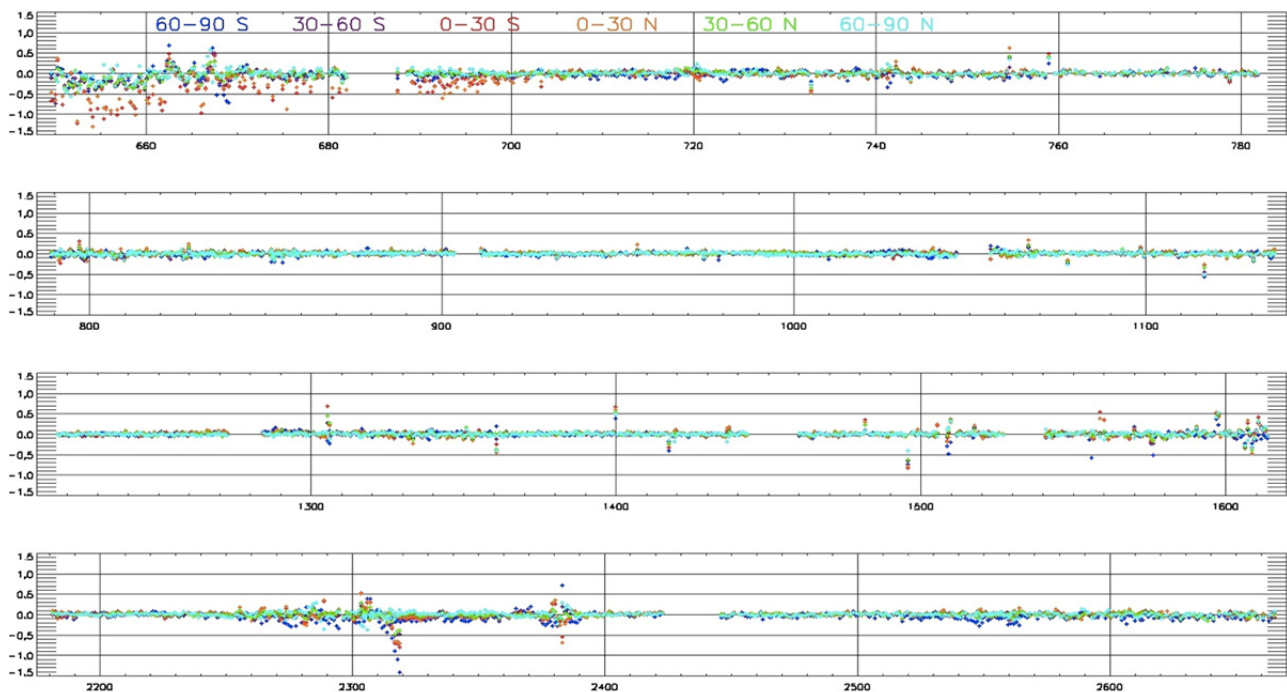


Figure 8. Bias of reconstruction error (K) as a function of frequency ( $\text{cm}^{-1}$ ) by latitude zone.

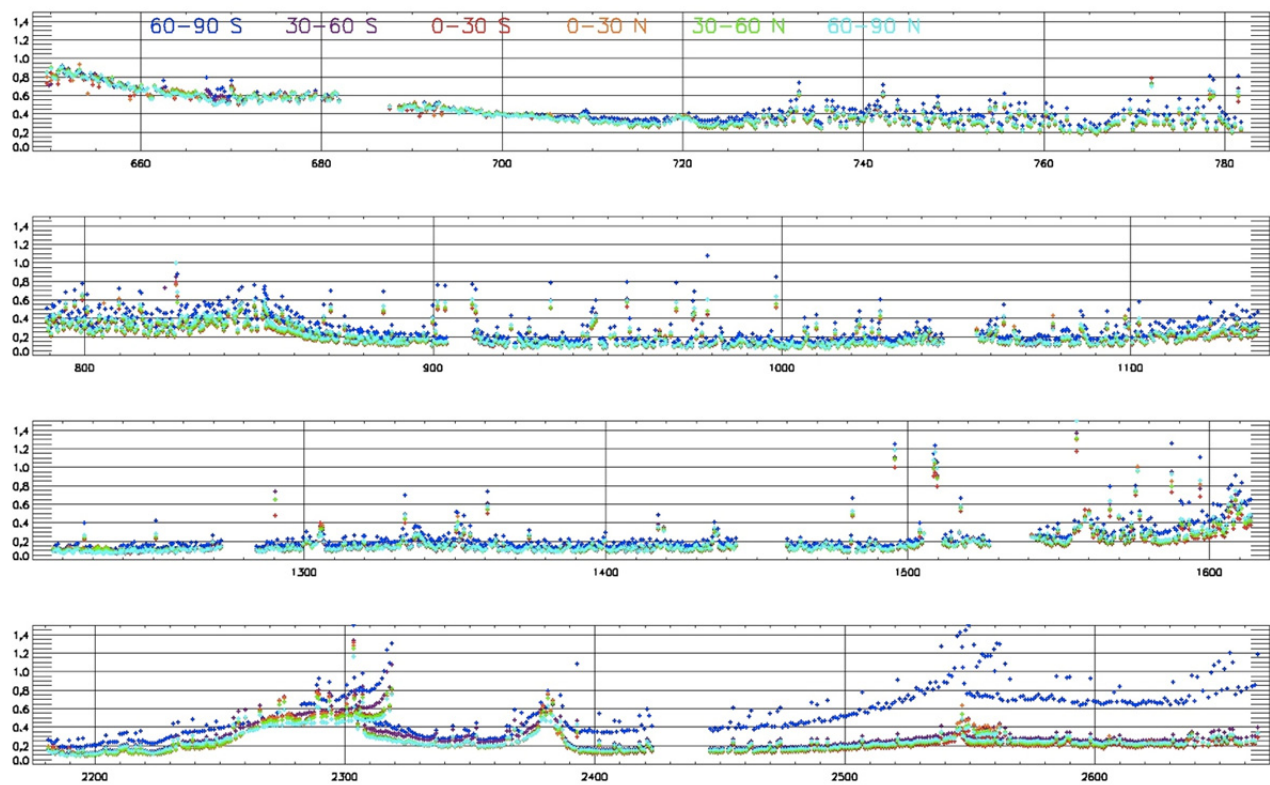


Figure 9. Standard deviation of reconstruction error (K) as a function of frequency ( $\text{cm}^{-1}$ ) by latitude zone.

We can also look at histograms of the bias and standard deviation of per-granule reconstruction error for the full day. Figure 10 shows that even at the granule level, outliers are extremely rare. There is evidence in Figure 10 of a small cold bias overall with a magnitude  $\ll 0.1$  K.

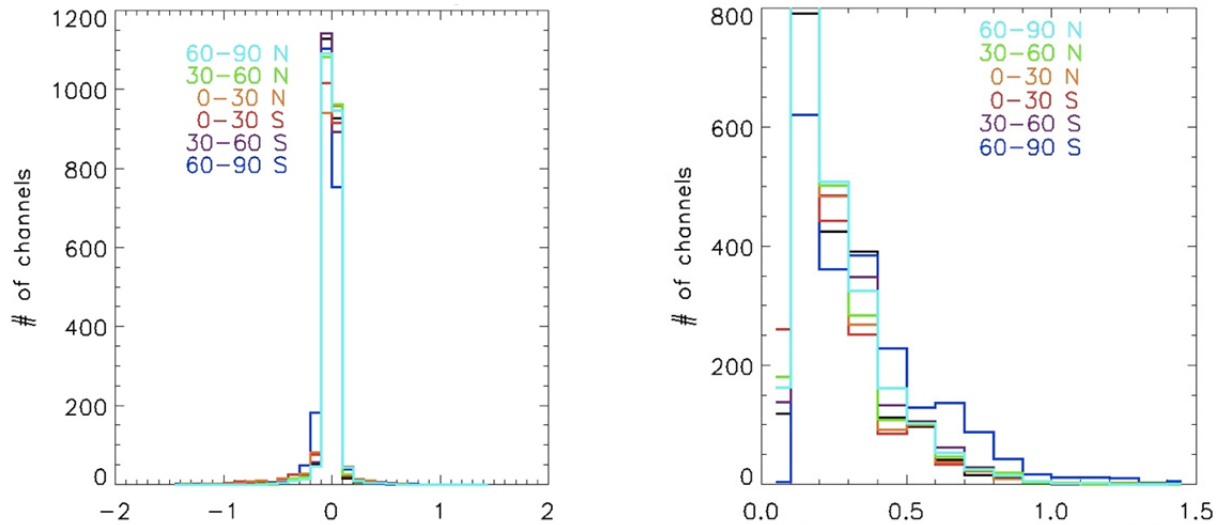


Figure 10 Left: histogram of fill error bias (K) by latitude zone; Right: histogram of fill error standard deviation (K) by latitude zone

## 2.7 Using PCA reconstructed radiances to explore the quality of raw Level-1B

PCA provides a great tool to investigate misbehaving channels, channels that are not replaced by the cleaning process detailed above, but show significant differences between the observed Level-1B radiances and the principal component reconstruction of the same channel. It may also (optionally) be used to dynamically flag and/or replace problem channels.

For this analysis we again processed all 240 granules of 2007-09-06. This time we produced an output that cleaned all channels with PCA reconstructed values. For each granule/channel combination we checked for cases where cleaning was not performed by the nominal Level-1C but the difference between the observed and reconstructed radiances were significant. The criterion for significance is that at least one of the  $135 \times 90 = 12150$  spectra has both a difference in BT units of at least 5 K and in radiance units of at least 5.7 times the noise level (5.7 sigma gives 1-in-100-million false positives assuming Gaussian noise. Over all 2378 channels \* 12150 spectra, that works out to one false positive per 3-4 granules.)

The problem of Antarctic winter cases mentioned previously is also an issue for this analysis. Since these cases do not help us understand noise characteristics of the instrument, all granules centered south of 60 degrees south latitude are excluded from the analysis below.

For each of the 240 granules, a quick-look graphic is made with images of BT difference between Level-1B observations and PCA reconstructions for those channels that are flagged for reconstruction differences. Figure 11 shows the report for granule 2. This typical granule has 17 of 2378 channels that meet the criteria for reporting.

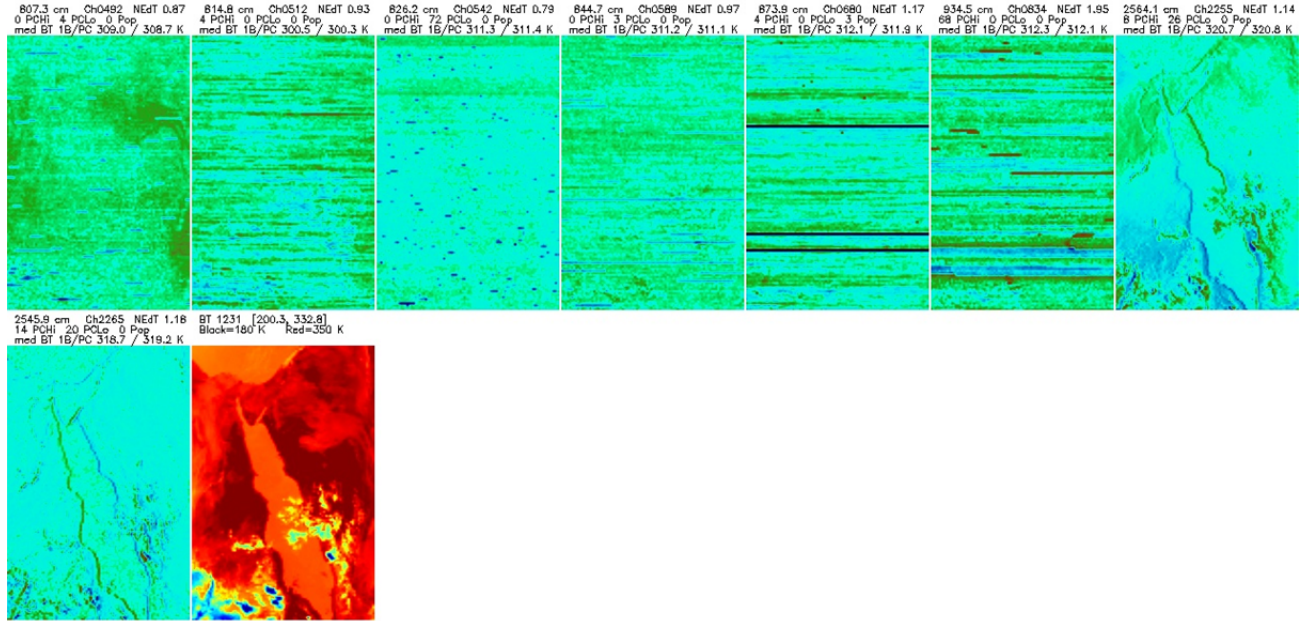


Figure 11 Per-channel images of differences between observed and reconstructed BTs, for all channels with at least one point of significant difference. Black stripes hide rows where the test is disabled because Level-1B detected a “pop” event.

The final red image shows the brightness temperature for this granule. The blue areas are cold clouds, and many shortwave channels are noisy in BT units in these areas. Because the test for significance includes a check in radiance units, these are probably not the differences that trigger these channels being flagged.

Many of the images have a strong horizontal component. Because AIRS operates in a whisk-broom fashion, short-duration events show up as short horizontal lines. Errors in calibration from noise during calibration observations get smoothed over  $\sim 10$  scans and so show up as more diffuse horizontal structures.

One of the primary non-Gaussian patterns encountered in AIRS data is “popping”, where the baseline dark current from a channel suddenly shifts and some time later shifts back to its original level. The Level-1B algorithm detects these events when they are large and they last for several scans. In these cases Level-1C always cleans the channel and these are shown as black lines in Figure 11. Most of the channels shown here have patterns that appear to be popping with short durations or small magnitude.

The most interesting case is channel 1579 ( $1399.9 \text{ cm}^{-1}$ ). It is a low noise channel ( $\text{NEdT}=0.12 \text{ K}$ ) where PCA and L1B agree very well except for 2 spectra in a row. This appears to be a radiation hit.

Except for the radiation hit case, all of the 17 channels featured in Figure 11 are known to have relatively high noise levels, but they were generally assumed to have Gaussian noise, so that they would be safe to use in climate applications where the noise will be reduced by averaging large numbers of data points. But since the noise now appears to be non-Gaussian, these channels should be avoided and Level-1C should clean them too.

Daily statistics show that these cases can be safely detected and flagged without significantly reducing the number of good channels. Figure 12 shows the noise levels of all channels, color coded by how often Level-1C already cleans them and how often the PCA test flags them as not matching expectations.



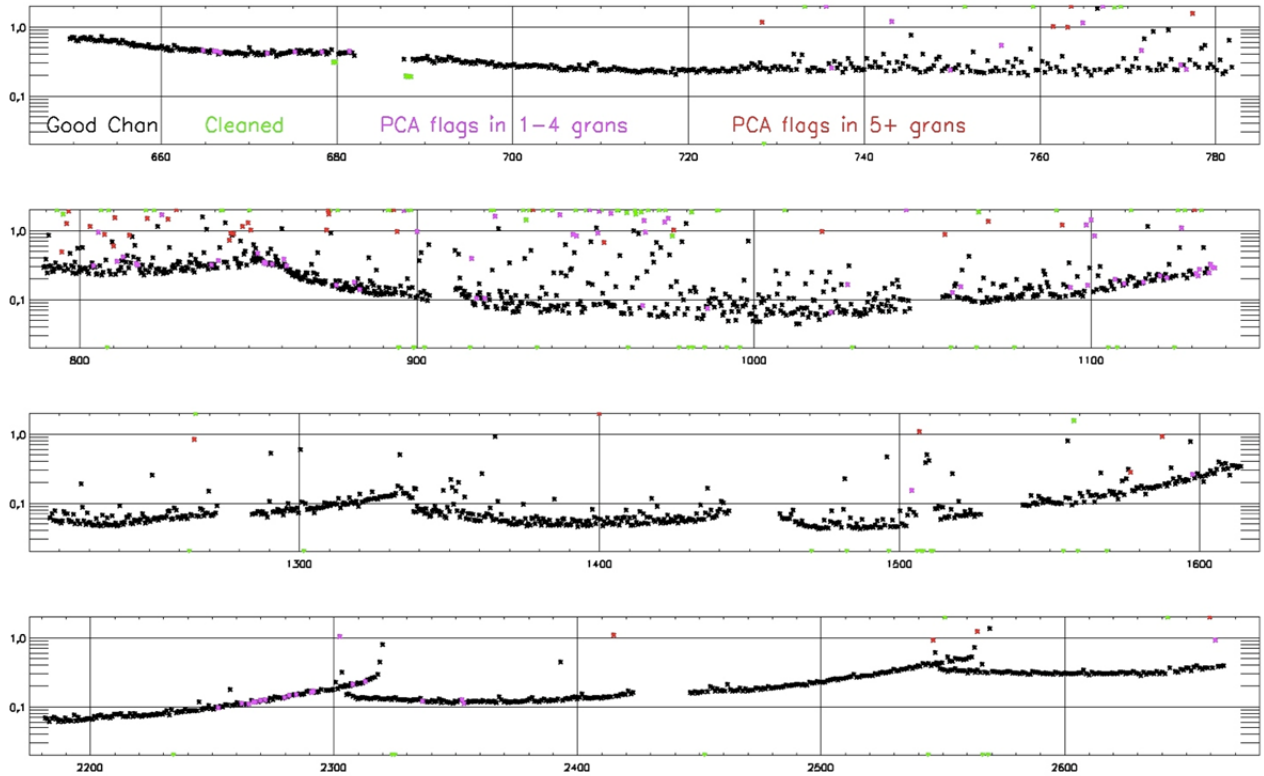


Figure 12 Daily maximum Noise Equivalent Delta Temperature (NEdT) (K) as a function of frequency ( $\text{cm}^{-1}$ ) for all AIRS channels. Channels where only one of two detectors is used have their NEdT reduced by  $\sqrt{2}$  to enhance the linear structure among the best channels. Colors show which channels were cleaned by Level-1C and which were not cleaned but had some  $5K/5.7 \times \text{NEN}$  outliers with respect to PCA reconstruction.

Figure 12 shows that all channels with 5 or more granules containing significant differences between L1B observed radiances and PCA reconstruction have noise levels significantly higher than is common in neighboring channels. So it should be possible to craft a model of expected noise level per channel, and mark as questionable and/or replace with reconstructed values any channel with  $2\times$  (TBD) this level of noise. This may be part of a future effort.

There are a number of channels colored magenta among the best channels. These had only a few (usually 1) events. These events are likely radiation hits, so for these cases just the few affected measurements should be flagged and/or replaced.

With the existing L1C cleaning channel criteria, typically about 5% of channels are flagged as bad and filled for each granule. Adding a criterion replacing all channels with noise greater than  $2\times$  the value for the best neighboring channels would flag and clean an additional 5-10%. Users then would have very high confidence in the quality of the remaining 85-90% pristine channels, and would also have high quality substitute values for the 10-15% flagged channels.

There are many channels with noise over  $2\times$  the local baseline that are colored black in Figure 12, indicating that no problems were detected with the PCA criteria. These could be good channels with slightly elevated but Gaussian noise, or they could have non-Gaussian noise at a magnitude that is not detected. More investigation will be needed before tighter cleaning rules are implemented operationally.

For the remaining channels, PCA could be used to flag and replace the very rare radiation hits.

## 2.8 Using PCA reconstructed radiances to explore spatial misalignment Level-1B

PCA reconstruction can also highlight spatial misalignment among channels. There are a small number of channels that see significantly more of one side or the other of an observed scene. The two most extreme among the channels that pass current quality screening are channels 2255 and 2265 at  $2564.1$  and  $2545.9$   $\text{cm}^{-1}$  respectively. These are the last good channels at the near ends of two adjacent detector modules, and have their observation centroids displaced opposite directions along the cross-track (roughly east-west) axis. Note that because both of these channels have noise levels over 1 K, both of these channels would be cleaned if we adopt the tighter noise tolerances discussed in section 2.7.

The PCA predicts with high accuracy what it expects an ideal channel at each frequency to see. The difference between what is actually observed and what PCA predicts functions as a sort of edge detection.

Figure 13a shows the brightness temperature at  $1231$   $\text{cm}^{-1}$  for 2007-09-06 granule #108. This is a daytime granule with a lot of hot desert, cooler sea (the Red Sea in the center and the Mediterranean Sea at top), and cold clouds. Figures 13b and 13c show the almost complementary PCA residuals for channels 2255 and 2265. Coastlines and edges of clouds are the most prominent features.

While these two channels have high enough noise levels that they would already be replaced according to the scheme proposed in section 2.7, this does point out that PCA could be used to clean up more subtle misalignment artifacts in less-affected channels. The differences between PCA and L1B observations for other channels rarely or never rise to the 5 K level of significance used in the section 2.7 analysis, but these other channels might be cleaned only for scenes where large, oppositely signed PCA residuals in channels 2255 and 2265 show that spatial misalignment is an issue.

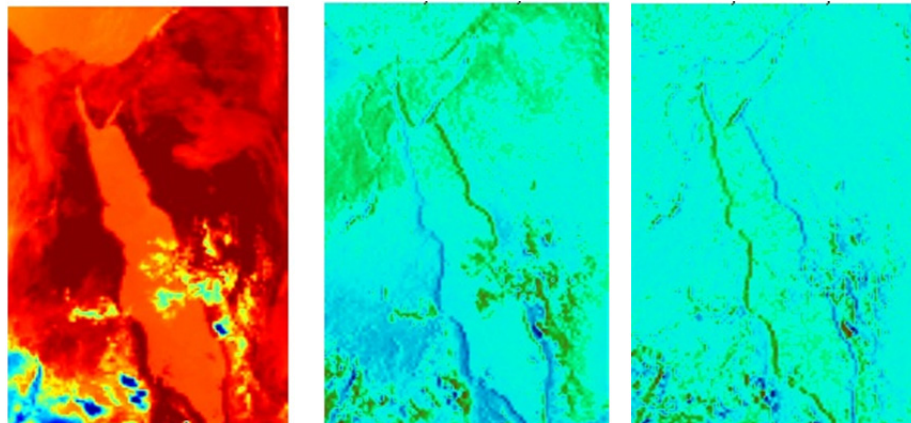


Figure 13 Daytime granule in Red Sea area. Left: False-color Brightness Temperature at  $1231$   $\text{cm}^{-1}$ ; Center: PCA residual for channel 2255; Right: PCA residual for channel 2265

## 3. CONCLUSIONS

The algorithm presented successfully corrects for the effect of time-variable non-Gaussian noise in AIRS spectra. This algorithm will be used as the basis for a new AIRS science data product—Level 1C—calibrated and cleaned radiances. That product will be made available to the public in a future version of the AIRS science software that runs at the GES DISC. Most channels have their calibrated radiances untouched from the basic product—Level 1B—calibrated radiances. A few channels have their radiances replaced by a sophisticated algorithm. This replacement process is performed independently for every spectrum. Replacement does not add significant bias and does not introduce undue noise. The Level 1C product is not a substitute for Level 1B (which will remain the primary AIRS Level 1 product), but an enhancement for optional use by users.



The PCA methodology has also shown promise as a tool for investigating features of marginal-quality AIRS channels, including non-Gaussian noise and spatial misalignment. The specific results shown here suggest that the final L1C should flag more channels for replacement, and that this can safely be done while leaving over 2000 channels unchanged from Level-1B.

## ACKNOWLEDGEMENTS

We thank Larrabee Strow and Scott Hannon of the University of Maryland Baltimore County (UMBC) for providing us with the model spectra used to determine replacement channels. This work was carried out at the Jet Propulsion Laboratory managed for NASA by the California Institute of Technology.

## REFERENCES

- [1] Strow, L. L., Scott E. Hannon, Margaret Weiler, Kenneth Overoye, Steven L. Gaiser, and Hartmut H. Aumann, Prelaunch Spectral Calibration of the Atmospheric Infrared Sounder (AIRS), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 41, NO. 2 (2003).
- [2] Gaiser, S. L., Hartmut H. Aumann, L. Larrabee Strow, Scott E. Hannon, and Margaret Weiler, In-Flight Spectral Calibration of the Atmospheric Infrared Sounder, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 41, NO. 2 (2003).
- [3] Denis A. Elliott, Hartmut H. Aumann, Yibo Jiang and Steven E. Broberg, "Level 1C spectra from the Atmospheric Infrared Sounder (AIRS)", Proc. SPIE 7081, 70810L (2008); <http://dx.doi.org/10.1117/12.795491>
- [4] Weiler, Margaret H., Overoye, Kenneth R., Stobie, James A., O'Sullivan, Paul B., Gaiser, Steven L., Broberg, Steven E., and Elliott, Denis A., "Performance of the Atmospheric Infrared Sounder (AIRS) in the Radiation Environment of Low-Earth Orbit", Proceedings of the SPIE 5882, 347-353 (2005).
- [5] Jolliffe, I. T., Principal Component Analysis. Springer-Verlag. pp. 487. doi:10.1007/b98835. ISBN 978-0-387-95442-4 (1986).